What is claimed is:

1. In a database system, a sampling method for constructing a data structure based on the contents of a database comprising:

a) gathering an initial sample of data from the database and creating a first data structure from said initial sample;

b) gathering a second sample of data from the database;

c) determining an initial sufficiency of the data gathered from the database that is based on a comparison of the first data structure and the second sample of data; and

d) forming a resultant data structure by gathering an additional sample of data from the database and using the additional amount of data to form the resultant data structure wherein the amount of data gathered in the additional sample is based on the initial sufficiency determination.

2. The method of claim 1 wherein the resultant data structure is formed based on data gathered in the initial sample, the second sample and the additional sample.

3. The method of claim 1 wherein the first and resultant data structures are histograms.

4. The method of claim 1 wherein the initial and second data samples are randomly retrieved block samples that form a first amount of data that is initially gathered and then divided in half to provide the initial and second data samples.

5. The method of claim 4 wherein the initial and second data samples are sorted and used to form two histograms.

6. The method of claim 5 wherein an error metric of the two histograms are formed by cross correlating the contents of the two histograms to determine the initial sufficiency.

7. The method of claim 6 wherein the initial and second data samples are further sub-divided to form sub-samples used to form other histograms of differing sample sizes that are cross correlated to find an error metric relating to said differing sample sizes.

8. The method of claim 6 wherein the initial and second data samples are further sub-divided to form additional sub-samples of smaller size that are used to form other histograms that are cross correlated for use in finding an error metric relating to sample sizes for use in determining a size of the additional sample of data to gather from the database.

9. The method of claim 4 additionally comprising estimating distinct values of an attribute of the initial and second samples by eliminating records from the blocks that are duplicated within a given block and estimating distinct values by categorizing attributes as rarely or frequently occurring within the database.

10. A computer readable medium for performing computer instructions to implement the method of claim 1.

11. A database system for constructing histograms based on sampling the contents of the database comprising:

a) a database management component that gathers block size data segments from the database which in aggregate form a first sample of data having a first size;

b) a histogram construction component that forms a first histogram from the first sample of data; and

c) a correlation component that determines an initial sufficiency of the first sample of data gathered from the database based on a comparison of the first histogram and data from the first sample of data;

d) wherein said database management component gathers an additional sample of data used by said histogram construction component in creating a resultant histogram and the size of the additional sample is based on the initial sufficiency determination.

12. The system of claim 11 wherein the resultant histogram is formed by the histogram construction component based on data gathered in the first sample of data and the additional data.

13. The system of claim 11 wherein the first sample of data and the additional sample of data are randomly retrieved block samples.

14. The system of claim 11 wherein histogram construction component sorts the data in said first sample of data as it constructs the first histogram.

15. The system of claim 11 wherein the correlation component determines an error metric by cross correlating the contents of the first histogram with other data in said first sample of data to determine the initial sufficiency.

16. The system of claim 15 wherein the first sample of data is sub-divided to form sub-samples used to form histograms of differing sizes that are cross correlated to find an error metric relating to said differing sample sizes.

17. The system of claim 15 wherein the first sample of data is sub-divided to form additional sub-samples of smaller size that are used to form other histograms that are cross correlated for use in finding an error metric relating to sample sizes for use in determining a size of the additional sample of data to gather from the database.

18. In a database system, a sampling method for constructing a histogram based on the contents of a database comprising:

a) gathering an initial sample of data from the database and creating a histogram from said initial sample;

b) gathering a second sample of data from the database for comparison with said first histogram;

c) determining an initial sufficiency of the data gathered from the database that is based on a comparison of the second sample with the first histogram; and

d) if the determination of initial sufficiency indicates the data in said initial and second samples is adequate to represent the database, combining the initial and second samples to form a resultant histogram, but if the determination of initial sufficiency indicates the initial and second samples are inadequate to represent the database, gathering an additional data sample to combine with the initial and second samples to form the resultant histogram wherein a size of the additional data sample is based on the initial sufficiency determination.

19. The method of claim 18 wherein the data is gathered in blocks from random storage locations within the database.

20. In a database system, a system for constructing a data structure based on the contents of a database comprising:

a) means for gathering an initial sample of data from the database and creating a first data structure from said initial sample;

b) means for determining an initial sufficiency of the data gathered from the database that is based on a comparison of the first data structure and other data in the initial sample not used to create the first data structure; and

c) means for forming a resultant data structure by gathering an additional sample of data from the database and using the additional amount of data to form the resultant data structure wherein the amount of data gathered in the additional sample is based on the initial sufficiency determination.

21. The system of claim 20 wherein the resultant data structure is formed based on data gathered in the initial sample and the additional sample.

22. The system of claim 21 wherein the first and resultant data structures are histograms.

23. The system of claim 20 wherein the initial data sample is made up of randomly retrieved block samples that form a first amount of data that is divided in half to provide data to form the data structure and data to cross correlate against the first data structure.

24. The system of claim 23 wherein the initial data samples is sorted and used to form two histograms.

25. The system of claim 24 wherein an error metric of the two histograms are formed by cross correlating the contents of the two histograms to determine the initial sufficiency.

26. The system of claim 25 wherein the initial data sample is further sub-divided to form sub-samples used to form other histograms of differing sample sizes that are cross correlated to find an error metric relating to said differing sample sizes.

27. The system of claim 26 wherein the initial and second data samples are further sub-divided to form additional sub-samples of smaller size that are used to form other histograms that are cross correlated for use in finding an error metric relating to sample sizes for use in determining a size of the additional sample of data to gather from the database.

28. The system of claim 24 additionally comprising means for estimating distinct values of an attribute of the initial and second samples by eliminating records from the blocks that are duplicated within a given block and estimating distinct values by categorizing attributes as rarely or frequently occurring within the database.

29. In a database system, a method for estimating distinct values of database attributes comprising:

a) gathering a plurality of block sized samples from the database;

b) organizing records gathered from the database into a first set of records where an attribute of a record is duplicated in different blocks and a second set of records wherein the attribute of a record is not duplicated in different blocks; and

c) estimating the number of distinct values of records in the database based on records in said first and second sets.

30. The method of claim 29 where records in a block are scanned to find attributes with duplicate values and where all records found to have a duplicate value for the attribute are collapsed into a representative record within the block.

31. The method of claim 29 wherein the samples are gathered from random locations in the database.

32. A computer readable medium for performing computer instructions to implement the method of claim 20.

33. A database system for determining database statistics comprising:

a) a database management component for gathering block size data segments from the database; and

b) an estimating component that organizing records gathered from the database into a first set of records where an attribute of a record is duplicated in different blocks and a second set of records wherein the attribute of a record is not duplicated in different blocks; and estimates the number of distinct values of records in the database based on records in said first and second sets.

34. In a database system, a method for estimating distinct values of database attributes comprising:

a) randomly gathering a plurality of block sized samples from the database;

b) modifying the contents of the samples by evaluating records in a block to find records having the same value for an attribute and collapsing all records found to have the same value for the attribute into a representative record within the block to provide modified block samples; and

c) estimating the number of distinct values of records in the database based on records in said modified block samples.